

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/825, 692 A
Source: IFW16
Date Processed by STIC: 04/25/2006

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 04/25/2006

PATENT APPLICATION: US/10/825,692A

TIME: 10:14:42

Input Set : A:\substitute Sequence Listing.txt

Output Set: N:\CRF4\04252006\J825692A.raw

3 <110> APPLICANT: Hotez, Peter
 4 Ashcom, James
 5 Bdamchian, Mahnaz
 6 Zhan, Bin
 7 Wang, Yan
 8 Hawdon, John
 9 Loukas, Alexander
 10 Williamson, Angela
 11 Jones, Brian
 12 Bethony, Jeffrey
 13 Goud, Gaddam
 14 Botazzi, Maria E.
 15 Mendez, Susana
 17 <120> TITLE OF INVENTION: Hookworm Vaccine
 19 <130> FILE REFERENCE: 03740007aa
 21 <140> CURRENT APPLICATION NUMBER: 10/825,692A
 22 <141> CURRENT FILING DATE: 2004-04-16
 24 <150> PRIOR APPLICATION NUMBER: US 60/329,533
 25 <151> PRIOR FILING DATE: 2001-10-17
 27 <150> PRIOR APPLICATION NUMBER: US 60/332,007
 28 <151> PRIOR FILING DATE: 2001-11-23
 30 <150> PRIOR APPLICATION NUMBER: US 60/375,404
 31 <151> PRIOR FILING DATE: 2002-04-26
 33 <150> PRIOR APPLICATION NUMBER: PCT US02/33106
 34 <151> PRIOR FILING DATE: 2002-10-17
 36 <160> NUMBER OF SEQ ID NOS: 116
 38 <170> SOFTWARE: PatentIn version 3.3
 40 <210> SEQ ID NO: 1
 41 <211> LENGTH: 1451
 42 <212> TYPE: DNA
 43 <213> ORGANISM: Necator americanus
 45 <400> SEQUENCE: 1

46 atgtttttctc ctgtagtcgt cagtgtggta ttcacaatcg ctttctgcaa tgcgtctcca 60
 48 gcaagagaca gcttcggctg ctctaacagt gggataactg acagcgaccg gcaagcggtc 120
 50 ctgcacttcc acaacaatgc tcgtcgacgg gttgcgaaag gccttgagga tagcaactcc 180
 52 ggcaaaactga atccagcgaa gaacatgtac aagctgtcat gggactgtgc aatggaacag 240
 54 cagcttcagg atgccatcca gtcattgccc agcggctttg ctgggattca aggtgttgcg 300
 56 cagaatacaa tgagctgggtc aagctctggt ggataccccg atccatcggt aaagatagaa 360
 58 ccaacgctct ccggctgggtg gagtgggtgcg aaaaagaacg gcgtaggccc ggacaacaaa 420
 60 tacaccgggtg gtggtctctt cgccttctct aacatgggtat actccgaaac gacgaaactt 480
 62 ggctgcgctt acaaggtttg cggcactaaa ctggcgggtt catgcatcta taatggagtc 540
 64 gggatcatca caaatcaacc tatgtgggag acaggtcagg cttgccagac aggagcagac 600
 66 tgctccactt acaagaactc aggtcgcgag gacggccttt gcacgaaggg accagatgta 660

Cpg-6)

RAW SEQUENCE LISTING

DATE: 04/25/2006

PATENT APPLICATION: US/10/825,692A

TIME: 10:14:42

Input Set : A:\substitute Sequence Listing.txt

Output Set: N:\CRF4\04252006\J825692A.raw

```

68 ccagaaacaa accagcagtg cccctcaaac accggaatga ctgattcagt cagagatact      720
70 ttcctatcgg tgcacaatga gttcagatcg agtggtgccc gaggtctgga acccgacgct      780
72 ctgggcgga atgcaccaa agcagctaaa atgctcaaga tgggtgatga ctgtgaagtg      840
74 gaagcatcgg ccatcagaca tggaaataaa tgcgtctatc aacattctca tgggtgaagac      900
76 agacctggac taggagaaaa catctacaaa actagtgtac tcaaattcga caagaacaaa      960
78 gcagccaagc aggcttcaca actctgggtgg aatgagttaa aagagtacgg cgtcggccca     1020
80 tccaacgtcc ttaccactgc gttatggaat agacccaaca tgcagattgg tcaactacacc     1080
82 cagatggcat gggacaccac ctacaaactt ggatgtgcag ttgttttctg caatgatttc     1140
84 acattcggcg tttgtcagta tgggccagga ggcaattaca tgggtcatgt catctacact     1200
86 atgggcccagc cgtgctctca gtgttcgcct ggtgctactt gcagcgtgac cgaaggcttg     1260
88 tgcagcgtct cttaatcagt caacaataaa tatcttacag tgatgttggt gcttacaat     1320
90 tgcttctttt ccaatagaaa taccaatgtc aacatcacga gtttctttaa attcatcact     1380
92 tccactacta ggggtgattt gaataaaatt tcatttcata aagcaattac atccgcaaaa     1440
94 aaaaaaaaaa a                                     1451

97 <210> SEQ ID NO: 2
98 <211> LENGTH: 424
99 <212> TYPE: PRT
100 <213> ORGANISM: Necator americanus
102 <400> SEQUENCE: 2
104 Met Phe Ser Pro Val Val Val Ser Val Val Phe Thr Ile Ala Phe Cys
105 1 5 10 15
108 Asn Ala Ser Pro Ala Arg Asp Ser Phe Gly Cys Ser Asn Ser Gly Ile
109 20 25 30
112 Thr Asp Ser Asp Arg Gln Ala Phe Leu Asp Phe His Asn Asn Ala Arg
113 35 40 45
116 Arg Arg Val Ala Lys Gly Leu Glu Asp Ser Asn Ser Gly Lys Leu Asn
117 50 55 60
120 Pro Ala Lys Asn Met Tyr Lys Leu Ser Trp Asp Cys Ala Met Glu Gln
121 65 70 75 80
124 Gln Leu Gln Asp Ala Ile Gln Ser Cys Pro Ser Gly Phe Ala Gly Ile
125 85 90 95
128 Gln Gly Val Ala Gln Asn Thr Met Ser Trp Ser Ser Ser Gly Gly Tyr
129 100 105 110
132 Pro Asp Pro Ser Val Lys Ile Glu Pro Thr Leu Ser Gly Trp Trp Ser
133 115 120 125
136 Gly Ala Lys Lys Asn Gly Val Gly Pro Asp Asn Lys Tyr Thr Gly Gly
137 130 135 140
140 Gly Leu Phe Ala Phe Ser Asn Met Val Tyr Ser Glu Thr Thr Lys Leu
141 145 150 155 160
144 Gly Cys Ala Tyr Lys Val Cys Gly Thr Lys Leu Ala Val Ser Cys Ile
145 165 170 175
148 Tyr Asn Gly Val Gly Tyr Ile Thr Asn Gln Pro Met Trp Glu Thr Gly
149 180 185 190
152 Gln Ala Cys Gln Thr Gly Ala Asp Cys Ser Thr Tyr Lys Asn Ser Gly
153 195 200 205
156 Cys Glu Asp Gly Leu Cys Thr Lys Gly Pro Asp Val Pro Glu Thr Asn
157 210 215 220
160 Gln Gln Cys Pro Ser Asn Thr Gly Met Thr Asp Ser Val Arg Asp Thr
161 225 230 235 240

```

RAW SEQUENCE LISTING

DATE: 04/25/2006

PATENT APPLICATION: US/10/825,692A

TIME: 10:14:42

Input Set : A:\substitute Sequence Listing.txt

Output Set: N:\CRF4\04252006\J825692A.raw

```

164 Phe Leu Ser Val His Asn Glu Phe Arg Ser Ser Val Ala Arg Gly Leu
165           245           250           255
168 Glu Pro Asp Ala Leu Gly Gly Asn Ala Pro Lys Ala Ala Lys Met Leu
169           260           265           270
172 Lys Met Val Tyr Asp Cys Glu Val Glu Ala Ser Ala Ile Arg His Gly
173           275           280           285
176 Asn Lys Cys Val Tyr Gln His Ser His Gly Glu Asp Arg Pro Gly Leu
177           290           295           300
180 Gly Glu Asn Ile Tyr Lys Thr Ser Val Leu Lys Phe Asp Lys Asn Lys
181 305           310           315           320
184 Ala Ala Lys Gln Ala Ser Gln Leu Trp Trp Asn Glu Leu Lys Glu Tyr
185           325           330           335
188 Gly Val Gly Pro Ser Asn Val Leu Thr Thr Ala Leu Trp Asn Arg Pro
189           340           345           350
192 Asn Met Gln Ile Gly His Tyr Thr Gln Met Ala Trp Asp Thr Thr Tyr
193           355           360           365
196 Lys Leu Gly Cys Ala Val Val Phe Cys Asn Asp Phe Thr Phe Gly Val
197           370           375           380
200 Cys Gln Tyr Gly Pro Gly Gly Asn Tyr Met Gly His Val Ile Tyr Thr
201 385           390           395           400
204 Met Gly Gln Pro Cys Ser Gln Cys Ser Pro Gly Ala Thr Cys Ser Val
205           405           410           415
208 Thr Glu Gly Leu Cys Ser Ala Pro
209           420

```

212 <210> SEQ ID NO: 3

213 <211> LENGTH: 1893

214 <212> TYPE: DNA

215 <213> ORGANISM: Necator americanus

217 <400> SEQUENCE: 3

```

218 ggtactgcag ggtttaatta cccaagtttg agaccaacag ccatgatttg gcgaacgtgg      60
220 caagttctcg tggttctgta tgcggcgctg tccattacag ttgtgaacgc ctataaacac      120
222 attagctccg atcacgttgt aaatacaaca ctgggtcaga ttcgaggagt accacagaat      180
224 ttcgaaggca aaaaagttac cgcttttctt ggtgtgccat atggtcaacc accgactggg      240
226 gaactacgat tcagcaatcc gaaaatgggt cagcggtggg aaggtataaa gaatgctaca      300
228 acaccggctc agccatgctt ccacttccct gacagtaaat ttaagggatt tcgtgggtca      360
230 gagatgtgga atccgaaagg aaatatgacc gaggattgct tgaatatgaa tatctgggtc      420
232 ccacacgatg ctgatgggtc cgtgattgta tggattttcg gaggcggctt cttcaccggg      480
234 tcaccatctt tagatgttta caacggtact gctctagcag ccaagaaacg taccattggt      540
236 gtgaacataa actatcgatt ggggtcccttc ggtttccttt atctcggtga tgattctcgt      600
238 gcacaaggga atatgggact gcaagatcaa caagttgcat tgcgatgggt gcataaacat      660
240 ataagctcct ttggtggaga tccgagaaaa gtcactcttt tcggcgaaag atcaggcgct      720
242 gcttcagcaa ccgctcatct agcagcaccg ggaagctatg agtttttcga taagataatt      780
244 ggcaacgggt gcacaatcat gaatagttgg gccagtcgaa caaatacatc gatgcttgag      840
246 ctgtcaatga aacttgctga acggttgaac tgtaccaaga aaagaaaaga cccgaatact      900
248 gtacatcgct gtttggttaa acatccagca catgtgggtc taaaagaggc cgctgttggt      960
250 tcgtatcaaa ttggtctcgt gctgacgttt gccttcatac ccattacctc tgataagaac      1020
252 ttcttcaggg gaaatgtctt tgatcgtcta cgagataaag acattaagaa gaatgtatcc      1080
254 attgtgcttg gtactgtaaa agacgaagca accttctttt taccctacta ctttggtcac      1140
256 aacggtttct ctttcaataa ctcattctta gcagatgggg aagaaaacag agcactcata      1200

```

RAW SEQUENCE LISTING

DATE: 04/25/2006

PATENT APPLICATION: US/10/825,692A

TIME: 10:14:42

Input Set : A:\substitute Sequence Listing.txt

Output Set: N:\CRF4\04252006\J825692A.raw

```

258 aatatatcac agtataatta tgcgatgaat gcaactgcgc catcacttga aagctcactg 1260
260 gaaccacttt tagaagctta taagaacgtt tcgacgcgaa aagaagaagg tgaaagatta 1320
262 cgcgatggtg ttggctcgatt catgggcgac tacttctata cctgcagcgt cattgatttc 1380
264 gctaatatcg tctcagacat tattaatgga tctttgtata tgtattactt tactaagagg 1440
266 tcagtggcaa atccttggcc agagtggatg ggtgtaatgc atgggttatga aatagaatac 1500
268 gaatttggac agcctttcct aaattcatca ctgtacaagg aaaagcttga aaacgaaaag 1560
270 atcttctcga aaaatatcat gagcttttgg aaagatttca tcaagactgg tgtccctgtc 1620
272 gatttttggc cgaaatacga tcgaaaggag cggaaagcgc tcgtacttgg cgaggaaagc 1680
274 gtgaacaatt cttaccctaa tatgactaat gttcatggac cgtactgtga actgatcgaa 1740
276 gaagcaaagg cgtctacaaa taatggactc accttgaaga aatacattga aggggagata 1800
278 aaaaataacg aaacgaacgt attttgatag aatgattttg cacagaatga agaattgaat 1860
280 atcaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 1893
283 <210> SEQ ID NO: 4
284 <211> LENGTH: 594
285 <212> TYPE: PRT
286 <213> ORGANISM: Necator americanus
288 <400> SEQUENCE: 4
290 Met Ile Trp Arg Thr Trp Gln Val Leu Val Val Leu Tyr Ala Ala Leu
291 1 5 10 15
294 Ser Ile Thr Val Val Asn Ala Tyr Lys His Ile Ser Ser Asp His Val
295 20 25 30
298 Val Asn Thr Thr Leu Gly Gln Ile Arg Gly Val Pro Gln Asn Phe Glu
299 35 40 45
302 Gly Lys Lys Val Thr Ala Phe Leu Gly Val Pro Tyr Gly Gln Pro Pro
303 50 55 60
306 Thr Gly Glu Leu Arg Phe Ser Asn Pro Lys Met Val Gln Arg Trp Glu
307 65 70 75 80
310 Gly Ile Lys Asn Ala Thr Thr Pro Ala Gln Pro Cys Phe His Phe Pro
311 85 90 95
314 Asp Ser Lys Phe Lys Gly Phe Arg Gly Ser Glu Met Trp Asn Pro Lys
315 100 105 110
318 Gly Asn Met Thr Glu Asp Cys Leu Asn Met Asn Ile Trp Val Pro His
319 115 120 125
322 Asp Ala Asp Gly Ser Val Ile Val Trp Ile Phe Gly Gly Gly Phe Phe
323 130 135 140
326 Thr Gly Ser Pro Ser Leu Asp Val Tyr Asn Gly Thr Ala Leu Ala Ala
327 145 150 155 160
330 Lys Lys Arg Thr Ile Val Val Asn Ile Asn Tyr Arg Leu Gly Pro Phe
331 165 170 175
334 Gly Phe Leu Tyr Leu Gly Asp Asp Ser Arg Ala Gln Gly Asn Met Gly
335 180 185 190
338 Leu Gln Asp Gln Gln Val Ala Leu Arg Trp Val His Lys His Ile Ser
339 195 200 205
342 Ser Phe Gly Gly Asp Pro Arg Lys Val Thr Leu Phe Gly Glu Ala Ser
343 210 215 220
346 Gly Ala Ala Ser Ala Thr Ala His Leu Ala Ala Pro Gly Ser Tyr Glu
347 225 230 235 240
350 Phe Phe Asp Lys Ile Ile Gly Asn Gly Gly Thr Ile Met Asn Ser Trp
351 245 250 255

```

RAW SEQUENCE LISTING

DATE: 04/25/2006

PATENT APPLICATION: US/10/825,692A

TIME: 10:14:42

Input Set : A:\substitute Sequence Listing.txt

Output Set: N:\CRF4\04252006\J825692A.raw

```

354 Ala Ser Arg Thr Asn Thr Ser Met Leu Glu Leu Ser Met Lys Leu Ala
355      260      265      270
358 Glu Arg Leu Asn Cys Thr Lys Lys Arg Lys Asp Pro Asn Thr Val His
359      275      280      285
362 Arg Cys Leu Val Lys His Pro Ala His Val Val Leu Lys Glu Ala Ala
363      290      295      300
366 Val Val Ser Tyr Gln Ile Gly Leu Val Leu Thr Phe Ala Phe Ile Pro
367 305      310      315      320
370 Ile Thr Ser Asp Lys Asn Phe Phe Gln Gly Asn Val Phe Asp Arg Leu
371      325      330      335
374 Arg Asp Lys Asp Ile Lys Lys Asn Val Ser Ile Val Leu Gly Thr Val
375      340      345      350
378 Lys Asp Glu Ala Thr Phe Phe Leu Pro Tyr Tyr Phe Gly His Asn Gly
379      355      360      365
382 Phe Ser Phe Asn Asn Ser Phe Leu Ala Asp Gly Glu Glu Asn Arg Ala
383      370      375      380
386 Leu Ile Asn Ile Ser Gln Tyr Asn Tyr Ala Met Asn Ala Thr Ala Pro
387 385      390      395      400
390 Ser Leu Glu Ser Ser Leu Glu Pro Leu Leu Glu Ala Tyr Lys Asn Val
391      405      410      415
394 Ser Thr Arg Lys Glu Glu Gly Glu Arg Leu Arg Asp Gly Val Gly Arg
395      420      425      430
398 Phe Met Gly Asp Tyr Phe Tyr Thr Cys Ser Val Ile Asp Phe Ala Asn
399      435      440      445
402 Ile Val Ser Asp Ile Ile Asn Gly Ser Leu Tyr Met Tyr Tyr Phe Thr
403      450      455      460
406 Lys Arg Ser Val Ala Asn Pro Trp Pro Glu Trp Met Gly Val Met His
407 465      470      475      480
410 Gly Tyr Glu Ile Glu Tyr Glu Phe Gly Gln Pro Phe Leu Asn Ser Ser
411      485      490      495
414 Leu Tyr Lys Glu Lys Leu Glu Asn Glu Lys Ile Phe Ser Lys Asn Ile
415      500      505      510
418 Met Ser Phe Trp Lys Asp Phe Ile Lys Thr Gly Val Pro Val Asp Phe
419      515      520      525
422 Trp Pro Lys Tyr Asp Arg Lys Glu Arg Lys Ala Leu Val Leu Gly Glu
423      530      535      540
426 Glu Ser Val Asn Asn Ser Tyr Pro Asn Met Thr Asn Val His Gly Pro
427 545      550      555      560
430 Tyr Cys Glu Leu Ile Glu Glu Ala Lys Ala Ser Thr Asn Asn Gly Leu
431      565      570      575
434 Thr Leu Lys Lys Tyr Ile Glu Gly Glu Ile Lys Asn Asn Glu Thr Asn
435      580      585      590
438 Val Phe
442 <210> SEQ ID NO: 5
443 <211> LENGTH: 1344
444 <212> TYPE: DNA
445 <213> ORGANISM: Necator americanus
447 <400> SEQUENCE: 5
448 ctcgtgccga attcggcacg agctccattc atcatgcage gatcattcct acttctactt

```

60

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 04/25/2006
PATENT APPLICATION: US/10/825,692A TIME: 10:14:43

Input Set : A:\substitute Sequence Listing.txt
Output Set: N:\CRF4\04252006\J825692A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:51; N Pos. 27,353,366,394,413

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:65,66,70,71,72,73,74,75,78,79,80,81,115,116

VERIFICATION SUMMARY

DATE: 04/25/2006

PATENT APPLICATION: US/10/825,692A

TIME: 10:14:43

Input Set : A:\substitute Sequence Listing.txt

Output Set: N:\CRF4\04252006\J825692A.raw

L:4146 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:51 after pos.:0
M:341 Repeated in SeqNo=51